

# Signal Reconstruction in Linear Mixing Systems with Different Error Metrics

Jin Tan and Dror Baron

Department of Electrical and Computer Engineering

North Carolina State University

Raleigh, NC 27695, USA

Email: {jtan,dzbaron}@ncsu.edu

**Abstract**—We consider the problem of reconstructing a signal from noisy measurements in linear mixing systems. The reconstruction performance is usually quantified by standard error metrics such as squared error, whereas we consider any additive error metric. Under the assumption that relaxed belief propagation (BP) can compute the posterior in the large system limit, we propose a simple, fast, and highly general algorithm that reconstructs the signal by minimizing the user-defined error metric. For two example metrics, we provide performance analysis and convincing numerical results. Finally, our algorithm can be adjusted to minimize the  $\ell_\infty$  error, which is not additive. Interestingly,  $\ell_\infty$  minimization only requires to apply a Wiener filter to the output of relaxed BP.

## I. INTRODUCTION

### A. Motivation

Linear mixing systems are popular models used in many settings such as compressed sensing [4, 5], regression [6, 7], and multiuser detection [8]. In this paper, we consider the following linear system [4, 5, 8],

$$\mathbf{w} = \Phi \mathbf{x}, \quad (1)$$

where the entries of the system input vector  $\mathbf{x} \in \mathbb{R}^N$  are independent and identically distributed (i.i.d.), the random linear mixing matrix (or measurement matrix)  $\Phi \in \mathbb{R}^{M \times N}$  is sparse and known, and the measurement vector is  $\mathbf{w} \in \mathbb{R}^M$ . Because each component in  $\mathbf{w}$  is a linear combination of the components in  $\mathbf{x}$ , we call the system (1) a *linear mixing system*.

The measurements  $\mathbf{w}$  are passed through a bank of separable scalar channels characterized by conditional distributions,

$$f_{\mathbf{Y}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) = \prod_{i=1}^M f_{Y_i|W_i}(y_i|w_i), \quad (2)$$

where  $\mathbf{y}$  is the channel output vector, and  $(\cdot)_i$  denotes the  $i$ th element of the corresponding vector. Note that the channels are general and are not restricted to Gaussian [9, 10]. Our goal is to reconstruct the original system input  $\mathbf{x}$  from the channel output  $\mathbf{y}$  and the measurement matrix  $\Phi$ .

This work was supported by the National Science Foundation, grant no. CCF-1217749, and by the U.S. Army Research Office, grant no. W911NF-04-D-0003. Portions appeared at the IEEE Statistical Signal Processing workshop (SSP), Aug. 2012 [1], and additional manuscripts [2, 3].

### B. Reconstruction quality

The performance of the estimation process is often characterized by some additive error metric that quantifies the distance between the estimated signal and the original signal. For a signal  $\mathbf{x}$  and its estimate  $\hat{\mathbf{x}}$ , both of length  $N$ , the error between them is the summation over the component-wise errors,

$$D(\hat{\mathbf{x}}, \mathbf{x}) = \sum_{j=1}^N d(\hat{x}_j, x_j). \quad (3)$$

Squared error, i.e.,  $d(\hat{x}_j, x_j) = |\hat{x}_j - x_j|^2$  in (3), is one of the most popular error metrics in various problems, due to many of its mathematical advantages. For example, minimum mean squared error (MMSE) estimation provides both variance and bias information about an estimator [11], and in the Gaussian case it is linear and thus often easy to implement [12]. However, there are applications where MMSE estimation is inappropriate, for example it is sensitive to outliers [13, 14]. Therefore, alternative error metrics, such as mean absolute error (median), mean cubic error, or Hamming distance, are used instead. Considering the significance of various types of error metrics other than squared error, *a general estimation algorithm that can minimize any desired error metric is of interest.*

The error metric function (3) has an additive structure, and this particular structure makes it convenient to propose a scalar-estimation based algorithm. However, if we want to prevent any significant absolute errors, i.e., to minimize the  $\ell_\infty$  norm error,

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_\infty = \max_{j \in \{1, \dots, N\}} |\hat{x}_j - x_j|, \quad (4)$$

where the error is not additive, then *is it possible to utilize the additive error metric (3) to approximate a minimum mean  $\ell_\infty$  error estimator?*

### C. Related work

Squared error is most commonly used as the error metric in estimation problems given by (1) and (2). Mean-square optimal analysis and algorithms were introduced in [15–19] to estimate a signal from measurements corrupted by Gaussian noise; in [9, 10, 20], further discussions were made about the

circumstances where the output channel is arbitrary, while, again, the MMSE estimator was put forth. Another line of work, based on a greedy algorithm called *orthogonal matching pursuit*, was presented in [21,22] where the mean squared error decreases over iterations. Absolute error is also under intense study in signal estimation. For example, an efficient sparse recovery scheme that minimizes the absolute error was provided in [23,24]; in [25], a fundamental analysis was offered on the minimal number of measurements required while keeping the estimation error within a certain range, and absolute error was one of the metrics concerned. Support recovery error is another metric of importance, for example it relates to properties of the measurement matrices [26]. The authors of [26–28] discussed the support error rate when recovering a sparse signal from its noisy measurements; support-related performance metrics were applied in the derivations of theoretical limits on the sampling rate for signal recovery [29, 30].

There have also been a number of studies on general properties of  $\ell_\infty$  error related solutions. An overdetermined linear system  $\mathbf{y} = \Phi\mathbf{x}$ , where  $\Phi \in \mathbb{R}^{N \times M}$  and  $N > M$ , was considered by Cadzow [31], and the properties of the minimum  $\ell_\infty$  error solutions to this system was explored. In Clark [32], the author developed a way to calculate the distribution of the greatest element in a finite set of random variables. And in Indyk [33], an algorithm was introduced to find the nearest neighbor of a point while  $\ell_\infty$  norm distance was considered. Finally, Lounici [34] studied the  $\ell_\infty$  error convergence rate for Lasso and Dantzig estimators.

#### D. Contributions

In this paper, we review our recent work on signal estimation with arbitrary additive error metrics [1, 2], and discuss its extension to minimizing the  $\ell_\infty$  norm error. Our contributions are: (i) we suggest a Bayesian estimation algorithm that minimizes an arbitrary additive error metric; (ii) we prove that the algorithm is optimal, and study the fundamental information-theoretic performance limit of estimation for a given metric; and (iii) we derive the performance limits for minimum mean absolute error and minimum mean support error.

The proposed metric-optimal algorithm is based on the assumption that the relaxed belief propagation (BP) method [10] converges to a set of degraded scalar Gaussian channels [9, 15, 17, 18]. The relaxed BP method is well-known to be optimal for the squared error, while we further show that the relaxed BP method can do more – because the sufficient statistics are given, other additive error metrics can also be minimized with one extra step, which is simple and fast. This is convenient for users who desire to recover the original signal with a non-standard additive error metric.

Another contribution that is included in this paper is an application of the metric-optimal algorithm to the  $\ell_\infty$  norm error, where the input signal is *sparse Gaussian* distributed, meaning that with probability  $s$  the vector entries are standard Gaussian, and with probability  $(1 - s)$  they are zero. We find [3] that, in the large system limit, the minimum mean  $\ell_\infty$

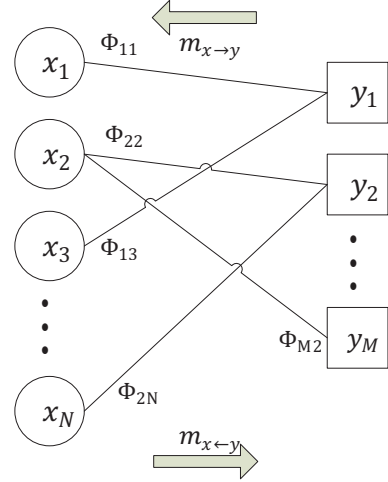


Fig. 1: Tanner graph for relaxed belief propagation.

estimator is a linear function of the output of the relaxed belief propagation (BP) method. Moreover, based on the metric-optimal algorithm, we propose a heuristic that achieves a lower  $\ell_\infty$  error than the linear function does when the signal dimension is finite.

The remainder of the paper is arranged as follows. We review the relaxed BP method in Section II. Then we describe our proposed algorithm, discuss its performance, and provide several example error metrics in Section III. We further discuss the possibility of extending the proposed algorithm for  $\ell_\infty$  norm error in Section IV. Simulation results appear in Section V.

## II. REVIEW OF RELAXED BELIEF PROPAGATION

Belief Propagation (BP) [35] is an iterative method used to compute the marginal probabilities of a Bayesian network. Consider the bipartite graph, called a *Tanner* or *factor* graph, shown in Figure 1, where circles represent random variables (called *variable nodes*), and related variables are connected through functions (represented by squares, called *factor nodes* or *function nodes*) that indicate their Bayesian dependence. In standard BP, there are two types of messages passed through the nodes: messages from variable nodes to factor nodes,  $m_{x \rightarrow y}$ , and messages from factor nodes to variable nodes,  $m_{y \rightarrow x}$ . If we denote the set of function nodes connected to the variable  $x$  by  $N(x)$ , the set of variable nodes connected to the function  $y$  by  $N(y)$ , and the factor function at node  $y$  by  $\Psi_y$ , then the two types of messages are defined as follows [35]:

$$m_{x \rightarrow y} = \prod_{k \in N(x) \setminus y} m_{k \rightarrow x},$$

$$m_{y \rightarrow x} = \sum_{\ell \in N(y) \setminus x} \Psi_y m_{\ell \rightarrow y},$$

where  $A \setminus B$  denotes the set whose elements are in  $A$  but not in  $B$ .

Inspired by the basic BP idea described above, the authors of [9, 16] developed iterative algorithms for estimation prob-

lems in linear mixing systems. In the Tanner graph, an input vector  $\mathbf{x} = [x_1, x_2, \dots, x_N]$  is associated with the variable nodes (input nodes), and the output vector  $\mathbf{y} = [y_1, y_2, \dots, y_M]$  is associated with the function nodes (output nodes). If  $\Phi_{ij} \neq 0$ , then nodes  $x_j$  and  $y_i$  are connected by an edge  $(i, j)$ , where the set of such edges  $E$  is defined as  $E = \{(i, j) : \Phi_{ij} \neq 0\}$ .

In standard BP methods [8, 36, 37], the distribution functions of  $x_j$  and  $w_i$  as well as the channel distribution function  $f_{Y|W}(y_i|w_i)$  were set to be the messages passed along the graph, but it is difficult to compute those distributions, making the standard BP method computationally expensive. In [16], a simplified algorithm, called *relaxed BP*, was suggested. In this algorithm, means and variances replace the distribution functions themselves and serve as the messages passed through nodes in the Tanner graph, greatly reducing the computational complexity. In [9, 10, 20, 38], this method was extended to a more general case where the channel is not necessarily Gaussian.

In Rangan [10], the relaxed BP algorithm generates two sequences,  $q_j(t)$  and  $\mu_j(t)$ , where  $t \in \mathbb{Z}^+$  denotes the iteration number. Under the assumptions that the signal dimension  $N \rightarrow \infty$  and the iteration number  $t \rightarrow \infty$ , while the ratio  $M/N$  is fixed, the sequences  $q_j(t)$  and  $\mu_j(t)$  converge to sufficient statistics for the linear mixing channel observation  $\mathbf{y}$  (2). More specifically, the conditional distribution  $f(x_j|q_j(t), \mu_j(t))$  converges to the conditional distribution  $f(x_j|\mathbf{y})$ , where  $q_j(t)$  can be regarded as a Gaussian-noise-corrupted version of  $x_j$ , and  $\mu_j(t)$  is the noise variance,

$$q_j(t) = x_j + v_j(t), \quad (5)$$

where  $v_j(t) \sim \mathcal{N}(0, \mu_j(t))$  for  $j = 1, 2, \dots, N$ . It has been shown [10] that  $\mu_j(t)$  converges to a fixed point that satisfies Tanaka's equation, which is analyzed in detail in [15, 17, 18, 20, 37, 39, 40]. We define the limits of the sequences,

$$\begin{aligned} \lim_{t \rightarrow \infty} q_j(t) &= q_j, \\ \lim_{t \rightarrow \infty} v_j(t) &= v_j, \\ \lim_{t \rightarrow \infty} \mu_j(t) &= \mu, \end{aligned}$$

for  $j = 1, 2, \dots, N$ . Note that all the scalar Gaussian channels (5) converge to the same noise variance  $\mu$ . We now simplify equation (5) as,

$$q_j = x_j + v_j, \quad (6)$$

where  $v_j \sim \mathcal{N}(0, \mu)$  for  $j = 1, 2, \dots, N$ .

### III. ESTIMATION ALGORITHM FOR ADDITIVE ERRORS

#### A. Algorithm

In this paper, we utilize the outputs of the relaxed BP method by Rangan [10], in particular the software package "GAMP" [41]. Figure 2 illustrates the structure of our *metric-optimal estimation algorithm* (dashed box). The inputs of the algorithm are: (i) a distribution function  $f_{\mathbf{x}}(\mathbf{x})$ , the prior of the original input  $\mathbf{x}$ ; (ii) a vector  $\mathbf{q} = [q_1, q_2, \dots, q_N]$ , the outputs of the scalar Gaussian channels computed by relaxed BP [10];

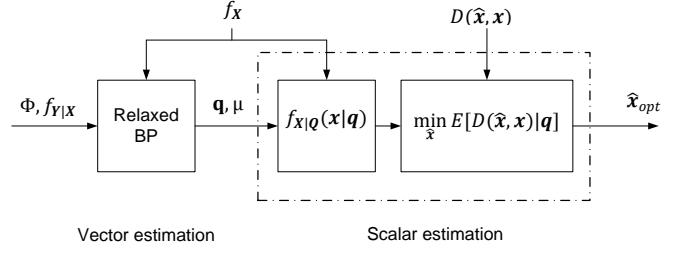


Fig. 2: The structure of the metric-optimal estimation algorithm.

(iii) a scalar  $\mu$ , the variance of the Gaussian noise in (6); and (iv) an additive error metric function  $D(\hat{\mathbf{x}}, \mathbf{x})$  (3) specified by the user.

Because the scalar channels have additive Gaussian noise, and the variances of the noise are all  $\mu$ , we can compute the conditional probability density function  $f_{\mathbf{x}|\mathbf{q}}(\mathbf{x}|\mathbf{q})$  from Bayes' rule:

$$\begin{aligned} f_{\mathbf{x}|\mathbf{q}}(\mathbf{x}|\mathbf{q}) &= \frac{f_{\mathbf{q}|\mathbf{x}}(\mathbf{q}|\mathbf{x})f_{\mathbf{x}}(\mathbf{x})}{f_{\mathbf{q}}(\mathbf{q})} \\ &= \frac{f_{\mathbf{q}|\mathbf{x}}(\mathbf{q}|\mathbf{x})f_{\mathbf{x}}(\mathbf{x})}{\int f_{\mathbf{q}|\mathbf{x}}(\mathbf{q}|\mathbf{x})f_{\mathbf{x}}(\mathbf{x})d\mathbf{x}}, \end{aligned} \quad (7)$$

where

$$f_{\mathbf{q}|\mathbf{x}}(\mathbf{q}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi\mu)^N}} \exp\left(-\frac{\|\mathbf{q} - \mathbf{x}\|_2^2}{2\mu}\right).$$

Given an error metric  $D(\hat{\mathbf{x}}, \mathbf{x})$ , the optimal estimand  $\hat{\mathbf{x}}_{\text{opt}}$  is generated by minimizing the conditional expectation of the error metric  $E[D(\hat{\mathbf{x}}, \mathbf{x})|\mathbf{q}]$ , which is easy to compute using  $f_{\mathbf{x}|\mathbf{q}}(\mathbf{x}|\mathbf{q})$ :

$$E[D(\hat{\mathbf{x}}, \mathbf{x})|\mathbf{q}] = \int D(\hat{\mathbf{x}}, \mathbf{x})f_{\mathbf{x}|\mathbf{q}}(\mathbf{x}|\mathbf{q})d\mathbf{x}.$$

Then,

$$\begin{aligned} \hat{\mathbf{x}}_{\text{opt}} &= \arg \min_{\hat{\mathbf{x}}} E[D(\hat{\mathbf{x}}, \mathbf{x})|\mathbf{q}] \\ &= \arg \min_{\hat{\mathbf{x}}} \int D(\hat{\mathbf{x}}, \mathbf{x})f_{\mathbf{x}|\mathbf{q}}(\mathbf{x}|\mathbf{q})d\mathbf{x}. \end{aligned} \quad (8)$$

The conditional probability  $f_{\mathbf{x}|\mathbf{q}}(\mathbf{x}|\mathbf{q})$  is separable, because the parallel scalar Gaussian channels (6) are separable and  $f_{\mathbf{x}}(\mathbf{x})$  is i.i.d. Moreover, the error metric function  $D(\hat{\mathbf{x}}, \mathbf{x})$  (3) is also separable. Therefore, the problem reduces to scalar estimation [12],

$$\begin{aligned} \hat{x}_{\text{opt},j} &= \arg \min_{\hat{x}_j} E[d(\hat{x}_j, x_j)|q_j] \\ &= \arg \min_{\hat{x}_j} \int d(\hat{x}_j, x_j)f_{x|q}(x_j|q_j)dx_j, \end{aligned} \quad (9)$$

for  $j = 1, 2, \dots, N$ . Equation (9) minimizes a single-variable function. We have shown in detail [1, 2] how to perform this minimization in two example error metrics, and summarize the results in Section III-C.

$$\text{MMUE}(f_{\mathbf{X}}, \mu) = \int_{R(\mathbf{Q})} \int_{R(\mathbf{X})} D(\hat{\mathbf{x}}_{\text{opt}}, \mathbf{x}) \left( \frac{1}{\sqrt{(2\pi\mu)^N}} \exp\left(-\frac{\|\mathbf{q} - \mathbf{x}\|^2}{2\mu}\right) \right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} d\mathbf{q}. \quad (10)$$

$$\text{MMAE}(f_{\mathbf{X}}, \mu) = N \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{\hat{x}_{j, \text{MMAE}}} (-x_j) f_{X_j|Q_j}(x_j|q_j) dx_j + \int_{\hat{x}_{j, \text{MMAE}}}^{+\infty} x_j f_{X_j|Q_j}(x_j|q_j) dx_j \right) f_{Q_j}(q_j) dq_j. \quad (11)$$

### B. Theoretical results

Having described the algorithm, we now discuss its theoretical properties based on the replica method. Because the replica method has not been rigorously justified, our results are given as claims. More detailed discussions of the claims can be found in Tan et al. [2].

*Claim 1:* Given the system model (1), (2) and an error metric  $D(\hat{\mathbf{x}}, \mathbf{x})$  of additive form (3), as the signal dimension  $N \rightarrow \infty$  and the measurement ratio  $M/N$  is fixed, the optimal estimand of the input signal is given by

$$\hat{\mathbf{x}}_{\text{opt}} = \arg \min_{\hat{\mathbf{x}}} E[D(\hat{\mathbf{x}}, \mathbf{x})|\mathbf{q}],$$

where the vector entries  $\mathbf{q} = (q_1, q_2, \dots, q_N)$  are the outputs of the scalar Gaussian channels (6).

As we mentioned in Section II, the probability density function  $f_{X_j|Y}(x_j|y)$  is statistically equivalent to  $f_{X_j|Q_j}(x_j|q_j)$  in the large system limit. Under this assumption, once we know the value of  $\mu$ , estimating each  $x_j$  from all channel outputs  $\mathbf{y} = (y_1, y_2, \dots, y_M)$  is equivalent to estimating  $x_j$  from the corresponding scalar channel output  $q_j$ . The relaxed BP method [10] calculates the sufficient statistics  $q_j$  and  $\mu$ . Therefore, an estimator based on minimizing the conditional expectation of the error metric,  $E(D(\hat{\mathbf{x}}, \mathbf{x})|\mathbf{q})$ , gives an asymptotically optimal result.

*Claim 2:* With the optimal estimand  $\hat{\mathbf{x}}_{\text{opt}}$  determined by (8), the minimum mean user-defined error (MMUE) is given by (10), where  $R(\cdot)$  represents the range of a variable, and  $\mu$  is the variance of the noise of the scalar Gaussian channel (6).

### C. Examples

We include examples that have been worked out in detail in Tan et al. [1, 2].

1) *Absolute error:* Because the MMSE is the mean of the conditional distribution, the outliers in the set of data may corrupt the estimation, and in this case the minimum mean absolute error (MMAE) could be a good alternative. The minimum mean absolute error is given by (11), where  $x_i$  (respectively,  $q_i$ ) is the input (respectively, output) of the scalar Gaussian channel (6),  $\hat{x}_{i, \text{MMAE}}$  is such that  $\int_{\hat{x}_{i, \text{MMAE}}}^{+\infty} f_{X_i|Q_i}(x_i|q_i) dx_i = \frac{1}{2}$ , and  $f_{X_i|Q_i}(x_i|q_i)$  is a function of  $f_{X_i}(x_i)$  following (7).

2) *Support recovery error:* In some applications, correctly estimating the locations where the data has non-zero values is almost as important as estimating the exact values of the data; it is a standard model selection error criterion [26]. The process of estimating the non-zero locations is called *support*

*recovery*. Support recovery error is defined as follows, and this metric function is discrete,

$$d_{\text{support}}(\hat{x}_j, x_j) = \text{xor}(x_j, \hat{x}_j),$$

where xor is the standard logical operator.

We have shown that the minimum mean support error (MMSuE) estimator achieves

$$\begin{aligned} \text{MMSuE}(f_{\mathbf{X}}, \mu) &= N(1-s) \cdot \text{erfc}\left(\sqrt{\frac{\tau}{2\mu}}\right) \\ &+ Ns \cdot \text{erf}\left(\sqrt{\frac{\tau}{2(\sigma^2 + \mu)}}\right), \end{aligned} \quad (12)$$

where

$$\tau = 2 \cdot \frac{\sigma^2 + \mu}{\sigma^2/\mu} \ln \left( \frac{(1-s)\sqrt{\sigma^2/\mu + 1}}{s} \right).$$

### IV. THE MINIMUM MEAN $\ell_\infty$ ERROR ESTIMATOR

Having discussed the metric-optimal algorithm, which was designed for additive error metrics, we now discuss how to extend the algorithm to support  $\ell_\infty$  error.

The  $\ell_\infty$  error only considers the component with greatest absolute value (4), and the minimum mean  $\ell_\infty$  error estimator is defined as

$$\hat{\mathbf{x}}_\infty = \arg \min_{\hat{\mathbf{x}}} \mathbf{E}[\|\hat{\mathbf{x}} - \mathbf{x}\|_\infty | \mathbf{q}]. \quad (13)$$

For a scalar Gaussian channel (6), we know that the minimum mean squared error estimator is achieved by conditional expectation. If  $x_j \sim \mathcal{N}(0, 1)$ , then it is the estimator  $\frac{1}{1+\mu}\mathbf{q}$  that gives the minimum mean squared error, where  $\mu$  is the variance of the noise in the scalar Gaussian channels (6). This format  $\frac{1}{1+\mu}\mathbf{q}$  is regarded as the *Wiener filter* in signal processing [42]. Our ongoing work [3] indicates that, in the large system limit, the Wiener filter is also optimal for  $\ell_\infty$  error when the input signal is *sparse* Gaussian.

*Claim 3:* For a set of parallel scalar Gaussian channels (6), where the input signal is sparse Gaussian distributed, the minimum mean  $\ell_\infty$  error estimator  $\hat{\mathbf{x}}_\infty \in \mathbb{R}^N$  (13) is linear in the channel observation vector  $\mathbf{q}$  as  $N \rightarrow \infty$ , i.e.,

$$\lim_{N \rightarrow \infty} \hat{\mathbf{x}}_\infty = \frac{\mu_x}{\mu_x + \mu_v} \mathbf{q}. \quad (14)$$

The detailed proof of Claim 3 is in preparation [3].

Unfortunately, the Wiener filtering (14) is only optimal for  $\ell_\infty$  error in the large system limit, and when the signal dimension is finite, our metric-optimal algorithm in conjunction

with a specific error metric *outperforms* the Wiener filtering in terms of  $\ell_\infty$  error.

Let us see what this specific error metric is. Recall that the definition of the  $\ell_p$  norm error between  $\hat{\mathbf{x}}$  and  $\mathbf{x}$  is

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_p = \left( \sum_{j \in \{1, \dots, N\}} |\hat{x}_j - x_j|^p \right)^{1/p},$$

and it can be shown that

$$\lim_{p \rightarrow +\infty} \|\hat{\mathbf{x}} - \mathbf{x}\|_p = \|\hat{\mathbf{x}} - \mathbf{x}\|_\infty.$$

It is therefore natural to turn to the  $\ell_p$  norm error as an alternative.

The  $\ell_p$  error is closely related to our definition of an additive error metric (3). We define

$$D_p(\hat{\mathbf{x}}, \mathbf{x}) = \sum_{j=1}^N |\hat{x}_j - x_j|^p = \|\hat{\mathbf{x}} - \mathbf{x}\|_p^p,$$

and let  $\hat{\mathbf{x}}_p$  denote the estimand that minimizes the conditional expectation of  $D_p(\hat{\mathbf{x}}, \mathbf{x})$ , i.e.,

$$\begin{aligned} \hat{\mathbf{x}}_p &= \arg \min_{\hat{\mathbf{x}}} E[D_p(\hat{\mathbf{x}}, \mathbf{x}) | \mathbf{q}] \\ &= \arg \min_{\hat{\mathbf{x}}} E[\|\hat{\mathbf{x}} - \mathbf{x}\|_p^p | \mathbf{q}], \end{aligned} \quad (15)$$

and

$$\hat{x}_{p,j} = \arg \min_{\hat{x}_j} E[|\hat{x}_j - x_j|^p | q_j],$$

for  $j \in \{1, 2, \dots, N\}$ .

The numerical results in Section V show that, for some values of  $p$ , the estimator  $\hat{\mathbf{x}}_p$  (15) achieves a lower  $\ell_\infty$  error than the Wiener filter (14).

## V. NUMERICAL RESULTS

To illustrate the performance of our algorithms, we provide some numerical results in this section, for both additive error metrics and the  $\ell_\infty$  norm error. The Matlab implementation of our algorithm can be found at [http://people.engr.ncsu.edu/dzbaron/software/arb\\_metric/](http://people.engr.ncsu.edu/dzbaron/software/arb_metric/). It automatically computes equations (7)-(9) where the additive error metric (3) is given as the input of the algorithm.

### A. Additive error metrics

We test our estimation algorithm on two cases modeled by (1) and (2): (i) sparse Gaussian input and Gaussian channel; (ii) sparse Weibull (defined below) input and Poisson channel. In both cases, the input's length  $N$  is 10,000, and its sparsity rate is  $s = 3\%$ , meaning that the entries of the input vector are non-zero with probability 3%, and zero otherwise. The matrix  $\Phi$  we use is Bernoulli(0.5) distributed, and is normalized to have unit-norm rows. In the first case, the non-zero input entries are  $\mathcal{N}(0, 1)$  distributed, and the Gaussian noise is  $\mathcal{N}(0, 3 \cdot 10^{-4})$  distributed, i.e., the signal-to-noise ratio

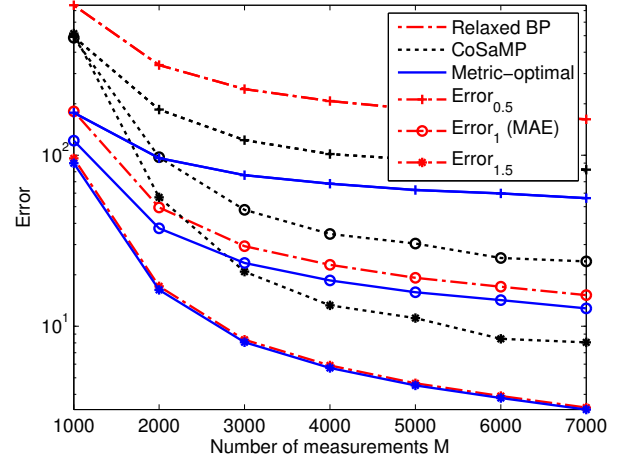


Fig. 3: Comparison of the metric-optimal estimation algorithm, relaxed BP, and CoSaMP. (Sparse Gaussian input and Gaussian channel; sparsity rate  $s = 3\%$ ; input length  $N=10,000$ ; SNR=20dB.)

(SNR) is 20dB. In the second case, the non-zero input entries are Weibull distributed,

$$f(x_j; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x_j}{\lambda}\right)^{k-1} e^{-(x_j/\lambda)^k} & x_j \geq 0 \\ 0 & x_j < 0 \end{cases},$$

where  $\lambda = 1$  and  $k = 0.5$ . The Poisson channel is

$$f_{Y|W}(y_i | w_i) = \frac{(\alpha w_i)^{y_i} e^{-(\alpha w_i)}}{y_i!}, \quad i \in \{1, 2, \dots, M\},$$

where the scaling factor of the input is  $\alpha = 100$ .

In order to illustrate that our estimation algorithm is suitable for reasonable error metrics, we considered absolute error and two other non-standard metrics:

$$\text{Error}_p = \sum_{j=1}^N |\hat{x}_j - x_j|^p,$$

where  $p = 0.5$  or  $1.5$ .

We compare our algorithm with the relaxed BP [10] and CoSaMP [22] algorithms. In Figure 3 and Figure 4, lines marked with “metric-optimal” present the errors of our estimation algorithm, and lines marked with “Relaxed BP” (respectively, “CoSaMP”) show the errors of the relaxed BP (respectively, CoSaMP) algorithm. Each point in the figure is an average of 100 experiments with the same parameters. Because the Poisson channel is not an additive noise channel and is not suitable for CoSaMP, the “MAE” and the “Error<sub>1.5</sub>” lines for “CoSaMP” in Figure 4 appear beyond the scope of the vertical axis. It can be seen that our metric-optimal algorithm outperforms the two other methods.

To demonstrate the theoretical analysis of our algorithm in Section III-C, we compare our MMAE estimation results with the theoretical limit (11) in Figure 5a, where the integrations are computed numerically. In Figure 5b, we compare our MMSuE estimator with the theoretical limit (12), where the value of  $\mu$  is acquired numerically from the relaxed BP

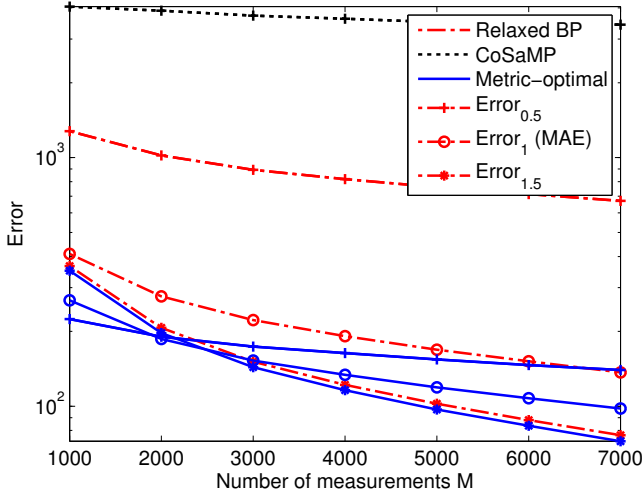


Fig. 4: Comparison of the metric-optimal estimation algorithm, relaxed BP, and CoSaMP. The “MAE” and the “Error<sub>1.5</sub>” lines for “CoSaMP” appear beyond the scope of the vertical axis. (Sparse Weibull input and Poisson channel; sparsity rate  $s = 3\%$ ; input length  $N=10,000$ ; input scaling factor  $\alpha = 100$ .)

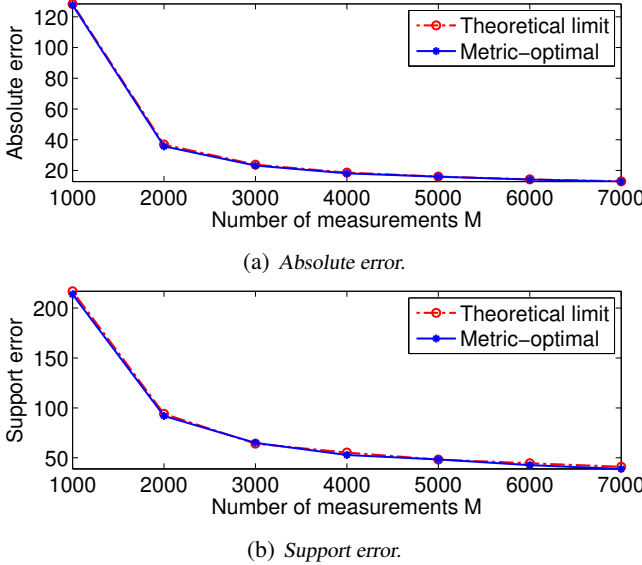


Fig. 5: Comparisons of the metric-optimal estimators and the corresponding theoretical limits (11), and (12). The corresponding lines coincide. (Sparse Gaussian input and Gaussian channel; sparsity rate  $s = 3\%$ ; input length  $N=10,000$ ; SNR=20dB.)

method [41] with 20 iterations. In these two figures, each point on the “metric-optimal” line is generated by averaging 40 experiments with the same parameters. It is shown in both figures that the lines coincide. Therefore our estimation algorithm reaches the corresponding theoretical limits and is optimal.

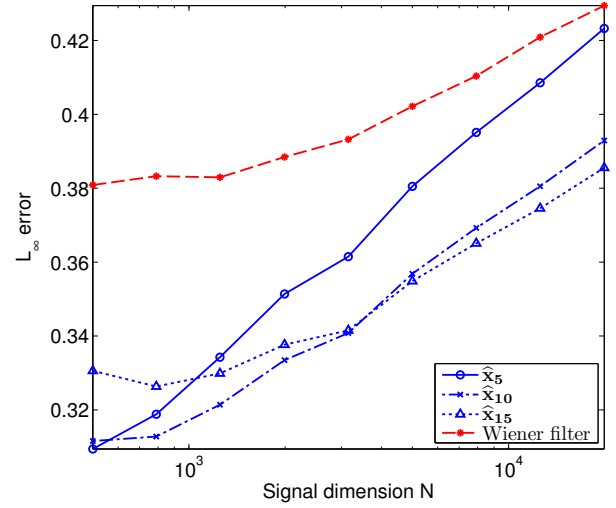


Fig. 6: The performance of the Wiener filter as well as the estimators  $\hat{x}_5$ ,  $\hat{x}_{10}$ , and  $\hat{x}_{15}$  in terms of  $\ell_\infty$  error. (Sparse Gaussian input and Gaussian channel; sparsity rate  $s = 5\%$ ; measurement ratio  $M/N = 0.3$ ; SNR=20dB.)

### B. The $\ell_\infty$ norm error

To show the performance of the Wiener filter and our metric-optimal algorithm in terms of the  $\ell_\infty$  error, we test the Wiener filter (14) as well as the estimator (15). The simulation settings remain the same with the previous simulations, except that we set the sparsity rate  $s = 5\%$ , and we only test the sparse Gaussian signal. The ratio  $M/N$  is fixed to 0.3, and the signal dimension  $N$  ranges from 500 to 20,000. We then run our metric-optimal algorithm with  $p = 5, 10, 15$  in (15). In Figure 6, we present the  $\ell_\infty$  norm error of the Wiener filter (14), along with the  $\ell_\infty$  norm error of  $\hat{x}_5$ ,  $\hat{x}_{10}$ , and  $\hat{x}_{15}$  (15). It can be seen that the estimators  $\hat{x}_5$ ,  $\hat{x}_{10}$ , and  $\hat{x}_{15}$  all achieve lower  $\ell_\infty$  errors than the Wiener filter (13) does, because the signal dimension is finite. Thus, for a finite signal dimension, it is reasonable to choose a proper value of  $p$ , and run the metric-optimal algorithm for  $\hat{x}_p$  (15) to achieve a low  $\ell_\infty$  error. That said, these results are specific to sparse Gaussian inputs, and it remains to be seen whether they will carry over to other input distributions.

### ACKNOWLEDGMENT

We thank Sundeep Rangan for kindly providing the Matlab code [41] of the relaxed BP algorithm [10, 20].

### REFERENCES

- [1] J. Tan, D. Carmon, and D. Baron, “Optimal estimation with arbitrary error metric in compressed sensing,” in *Proc. IEEE Stat. Sig. Proc. Workshop (SSP)*, Aug. 2012, pp. 588–591.
- [2] J. Tan, D. Carmon, and D. Baron, “Signal estimation with arbitrary error metrics in compressed sensing,” *arXiv:1207.1760*, July 2012.
- [3] J. Tan, D. Baron, and L. Dai, “Minimax error estimation in linear mixing systems,” in preparation.
- [4] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

- [5] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [6] P.J. Huber, "Robust regression: asymptotics, conjectures and Monte Carlo," *Ann. Stat.*, vol. 1, no. 5, pp. 799–821, 1973.
- [7] D.P. O'Leary, "Robust regression computation using iteratively reweighted least squares," *SIAM J. Matrix Anal. Appl.*, vol. 11, no. 3, pp. 466–480, July 1990.
- [8] D. Guo and C.-C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 421–431, Apr. 2008.
- [9] D. Guo and C.C. Wang, "Random sparse linear systems observed via arbitrary channels: A decoupling principle," in *Proc. Int. Symp. Inf. Theory (ISIT2007)*, June 2007, pp. 946–950.
- [10] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," *CoRR*, vol. arXiv:1001.2228v1, Jan. 2010.
- [11] U. Grenander and M. Rosenblatt, *Statistical analysis of stationary time series*, Wiley New York, 1957.
- [12] B.C. Levy, *Principles of signal detection and parameter estimation*, Springer Verlag, 2008.
- [13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 1991.
- [14] A.R. Webb, *Statistical pattern recognition*, John Wiley & Sons Inc., 2002.
- [15] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, June 2005.
- [16] D. Guo and C.C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *IEEE Inf. Theory Workshop*, Oct. 2006, pp. 194–198.
- [17] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. 47th Allerton Conf. Commun., Control, and Comput.*, Sep. 2009.
- [18] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *CoRR*, vol. abs/0906.3234, June 2009.
- [19] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, pp. 269–280, Jan. 2010.
- [20] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *Arxiv preprint arXiv:1010.5141*, Oct. 2010.
- [21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [22] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harm. Anal.*, vol. 26, no. 3, pp. 301–321, 2008.
- [23] P. Indyk and M. Ruzic, "Near-optimal sparse recovery in the  $\ell_1$  norm," in *49th Annu. IEEE Symp. Found. Comput. Sci.*, Oct. 2008, pp. 199–207.
- [24] R. Berinde, P. Indyk, and M. Ruzic, "Practical near-optimal sparse recovery in the  $\ell_1$  norm," in *Proc. 46th Allerton Conf. Commun., Control, and Comput.*, Sep. 2008, pp. 198–205.
- [25] A. Cohen, W. Dahmen, and R. A. DeVore, "Near optimal approximation of arbitrary vectors from highly incomplete measurements," 2007, preprint.
- [26] W. Wang, M.J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2967–2979, June 2010.
- [27] A. Tulino, G. Caire, S. Shamai, and S. Verdú, "Support recovery with sparsely sampled free random matrices," in *IEEE Int. Symp. Inf. Theory*, 2011, pp. 2328–2332.
- [28] M.J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.
- [29] M. Akçakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.
- [30] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Inform. Theory*, vol. 58, no. 5, pp. 3065–3092, May 2012.
- [31] J.A. Cadzow, "Minimum  $\ell_1$ ,  $\ell_2$ ,  $\ell_\infty$  norm approximate solutions to an overdetermined system of linear equations," *Digital Signal Proces.*, vol. 12, no. 4, pp. 524–560, Oct. 2002.
- [32] C.E. Clark, "The greatest of a finite set of random variables," *Oper. Res.*, vol. 9, no. 2, pp. 145–162, Mar. 1961.
- [33] P. Indyk, "On approximate nearest neighbors under  $\ell_\infty$  norm," *J. Comput. Syst. Sci.*, vol. 63, no. 4, pp. 627–638, Dec. 2001.
- [34] K. Lounici, "Sup-norm convergence rate and sign concentration property of lasso and Dantzig estimators," *Electron. J. Stat.*, vol. 2, pp. 90–102, 2008.
- [35] C.M. Bishop, *Pattern recognition and machine learning*, vol. 4, Springer New York, 2006.
- [36] G. Caire, R.R. Muller, and T. Tanaka, "Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation," *Information Theory, IEEE Transactions on*, vol. 50, no. 9, pp. 1950–1973, Sep. 2004.
- [37] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *IEEE Inf. Theory Workshop*, Mar. 2006, pp. 160–164.
- [38] S. Rangan, A.K. Fletcher, V.K. Goyal, and P. Schniter, "Hybrid approximate message passing with applications to structured sparsity," *Arxiv preprint arXiv:1111.2581*, Nov. 2011.
- [39] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
- [40] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.
- [41] S. Rangan, A. Fletcher, V. Goyal, U. Kamilov, J. Parker, and P. Schniter, "GAMP," [http://gampmatlab.wikia.com/wiki/Generalized\\_Approximate\\_Message\\_Passing/](http://gampmatlab.wikia.com/wiki/Generalized_Approximate_Message_Passing/).
- [42] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series with engineering applications*, MIT press, 1949.